

RECEIVED
CENTRAL FAX CENTER

;+33 0476614103

4/ 57

APR 23 2004

OFFICIAL

PATENT APPLICATION
Attorney Docket No. D/A0A34

CERTIFICATE OF TRANSMISSION

I hereby certify that this correspondence is being facsimile transmitted to the Patent and Trademark Office to Fax No.: (703) 872-9306 on April 23, 2004.

Typed or printed name of person signing this certificate
Thomas Zell

Signature:

Thomas Zell

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of: Gregory T. Grefenstette et al.)

Appl. No.: 09/683,235

Filed: December 5, 2001

Art Unit: 2172

Examiner: Isaac M. Woo

Title: SYSTEM FOR AUTOMATICALLY GENERATING QUERIES

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF

Sir:

Appellant respectfully submits this Appeal Brief in the appeal of the present case to the Board of Appeals and Patent Interferences on the Notice dated February 23, 2004.

05/05/2004 KWATSON 00000001 240025 09683235

01 FC:1401 330.00 DA

Appl. No. 09/683,235

1. Real Party of Interest

The real party of interest in the present application is the assignee of the present application, Xerox Corporation.

2. Related Appeals and Interferences

There is no related appeal or interference other than appeal briefs have been filed for co-pending co-assigned U.S. Patent Application Serial Nos. 09/683,237 entitled "System With User Directed Enrichment And Import/Export Control" and 09/683,242 "Document-Centric System With Auto-Completion And Auto-Correction", which was filed concurrently with the instant Application and similar to the instant Application claims priority to U.S. Provisional Application 60/311,857.

3. Status of the Claims

Claims 1-20 are pending in this application. Of these, claims 1, 14, and 18 are independent claims. An Amendment faxed September 8, 2003 amended claims 1, 8, 14, and 18. Claims 1-8 and 10-20 have been finally rejected in an Office Action mailed November 21, 2003 (hereinafter referred to as the "Office Action") with similar comments with regard thereto in an Advisory Action mailed February 9, 2004, on the grounds further discussed herein. The Office Action indicates that claim 9 is objected to but would be allowable if rewritten in independent form to include all of the limitations of the base claim and any intervening claims.

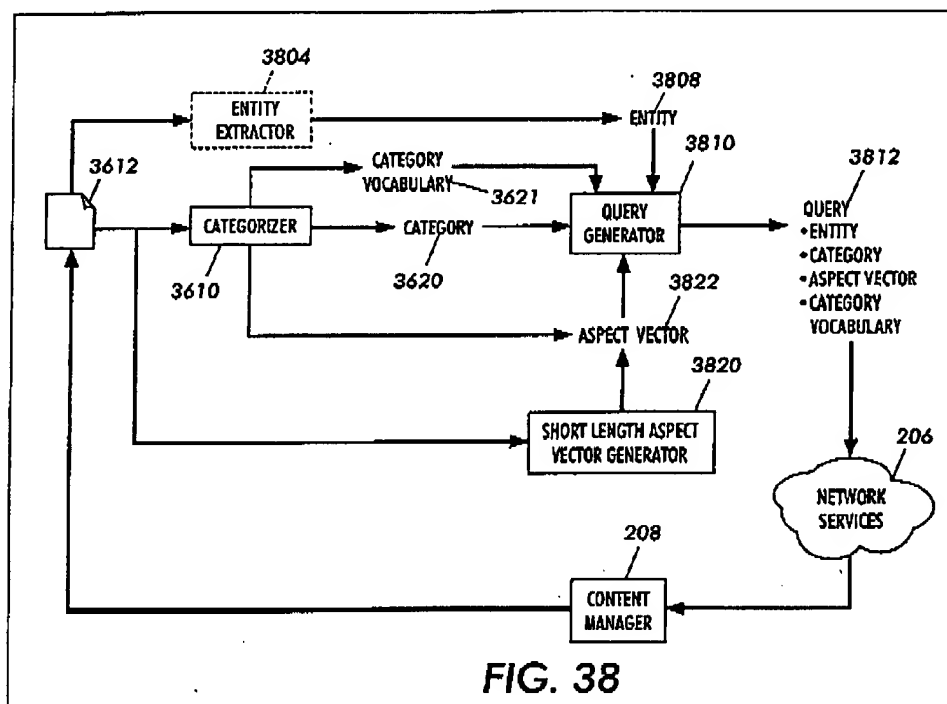
4. Status of Amendments

It is understood that all amendments to the claims made in this application have been entered and are reflected in the claims forming Appendix A hereto.

Appl. No. 09/683,235

5. Summary of Invention

Appellant's invention is directed at a method, system, and article of manufacture for automatically formulating a query, which is described in detail in section F.3 of Appellant's specification (see paragraph numbers 397-426). The system, as illustrated in Appellant's Figure 38 reproduced below, includes an entity extractor (3804), a categorizer (3610), and a query generator (3810). The entity extractor identifies a set of entities (3808) in selected document content (3612) for searching information related thereto in, for example, an information retrieval system (206). The categorizer defines an organized classification of content with each class in the organization having an associated classification label that corresponds to a category of information in the information retrieval system.



The categorizer assigns the selected document content a set of classification labels that defines a set of categories (3620) from the organized classification of content. The query generator automatically formulates a query (3812) concerning the set of entities extracted by the entity extractor. In formulating the query, the query generator restricts the search at the information retrieval system to the category of

Appl. No. 09/683,235

information in the information retrieval system identified by the assigned classification label.

In addition, the selected document content may be analyzed by a short length aspect generator (3820) to formulate a short run aspect vector (3822). Further, the categorizer may produce classification labels that identify a characteristic or category vocabulary (3621) associated with the corresponding classes. In one embodiment, the query generator coalesces these four elements (i.e., entity 3808, category 3620, aspect vector 3822, and category vocabulary 3621) to automatically formulate a query (3812). Results from the query may then be used by a content manager (208) to enrich the original document content (3612).

6. Issues

The single issue presented herein is whether claims 1-8 and 10-20 are unpatentable under 35 U.S.C. §103(a) over Rennison et al., U.S. Patent No. 6,154,213 (hereinafter referred to as "Rennison").

7. Grouping of Claims

The claims do not stand or fall together as a group and are grouped as follows:

FIRST GROUP: Independent claims 1, 14, and 18 and dependent claims 3, 5-8, and 10-13 define a first group of claims that for reasons discussed below stand or fall together.

SECOND GROUP: Claims 2, 15, and 19, which depend from claims 1, 14, and 18, respectively, define a second group of claims that for reasons discussed below stand or fall together.

THIRD GROUP: Claims 4, 16, and 20, which depend from dependent claims 2, 15, and 19, respectively, define a third group of claims that for reasons discussed below stand or fall together.

Claim 17, which depends from claim 16, stands on its own for reasons discussed below.

Appl. No. 09/683,235

8. Argument

Appellant traverses the final rejection of claims 1-8 and 10-20 under 35 U.S.C. §103(a) as being unpatentable over Rennison and submits for the reasons set forth below that Appellant's claimed invention is patentably distinguishable over Rennison.

A. Brief Summary of Rennison

Rennison discloses a method for navigating through large document collections by "maintaining a constant density of visual information presented on a display device to the user at any given moment in time" (see Rennison Abstract). The method disclosed by Rennison segments a large document collection into various units of information and provides "three different types of cues to the user: scale, context and an indication of the types of selected relationships between items of information in the information structure" (see Rennison column 3, lines 55-61).

More specifically, Rennison discloses that "the information structure of an information space is dynamically determined in response to a user's query and is a representation of the relationship between a collection of documents that satisfy the query" (see Rennison column 4, lines 43-47). Further, Rennison discloses that a user "creates queries by navigating through the 3D information space itself, which is dynamically repopulated with 3D graphical objects representing an information structure which is computed in response to the user's movements (query) in the 3D space" (see Rennison column 4, lines 57-61).

B. The First Group Of Claims Is Patentable Over Rennison

For the purpose of discussion presented herein, claim 1 is discussed as a representative claim of the first group, which includes independent claims 14 and 18. In rejecting the claims, the Office Action alleges that subject matter of the claimed invention is made obvious in view of the disclosure in columns 4-6, 8-10, 17-19, 21, and 26 of Rennison. Appellant respectfully disagrees.

Instead, Appellant respectfully submits that Rennison fails to disclose or suggest Appellant's claimed limitations set forth in claim 1 of: *automatically formulating a query to restrict a search at an information retrieval system for information concerning a set of*

Appl. No. 09/683,235

entities to a category of information in the information retrieval system identified by a classification label assigned by categorizing the selected document content.

B.1 Rennison Fails to Disclose or Suggest Restricting A Query To A Category Of Information In An Information Retrieval System Using A Classification Label

Rennison discloses a tool for navigating through a large document collection. Given a set of documents, the structuring process creates a "Space" (i.e., a graph as shown in Figure 1 of Rennison) "of Concepts that permits navigation of the set of documents" (see Rennison column 25, lines 11-15). The large document collections are developed during a "structuring process" which involves "recursively finding common Concepts that can group documents to provide coverage over a document set, and finding subtopics of these that provide distinction between these document to yield smaller document sets" (see Rennison column 25, lines 31-35).

In Rennison, the set of documents that is used to define the Space is identified through a user query (alternatively, a set of document is provided directly by the user) (see Rennison column 25, lines 12-13). (See also column 29, lines 14-15 – "the user query indicates what documents to build the graph around".) In contrast, Appellant claims a method for formulating a query using a document, not to use a query to define a set of documents that are used to build a graph.

Furthermore, Rennison expands terms of the user's query using a knowledge base (KB) (see Rennison column 26, lines 19-20). Rennison discloses that "all the KB Concepts which are related to or subsumed by the query term are also included in the search, so that it needs not rely on matching an exact word, but can instead match the general concept of interest" (see Rennison column 26, lines 32-35). (See also Rennison column 29, lines 16-32.)

Further as set forth in Rennison, "information retrieval and query formation are controlled by movement through the information space [] from one graphical node [] to another" (see column 12, lines 36-44). (See also Rennison column 13, lines 1-5 – "Thus, movement in the information space [] defines both the query to the information structure [], and the resulting display of the information space which is updated to reflect

Appl. No. 09/683,235

such movement".) That is, movement in the information space has the effect of defining a query.

Unlike either action described in Rennison (a) using user queries to define document making up an information space or (b) defining queries by user movement in the information space, Appellant's claimed invention concerns the automatic formulation of a query in which the automatically formulated query is *restricted* to information concerning a set of entities identified in document content and a category of information in an information retrieval system identified by a classification label assigned by categorizing the document content.

Moreover, Appellant does not dynamically generate an information space in response by expanding terms of a user query using a knowledge base as taught by Rennison, instead Appellant's claimed invention recited in independent claim 1 concerns the automatic formulation of a query from selected document content by, in part, (a) *categorizing* the selected document content, and (b) formulating a query to *restrict* a search to a category of information at an information retrieval system.

B.2 Rennison Fails to Disclose or Suggest Categorizing a Document to Formulate A Query

Besides the operation of identifying a set of documents that match a user's query, Rennison discloses another operation which involves building the information space using the set of identified documents (see Rennison generally from column 26, line 36 to column 28, line 61). The information space is built "by finding the smallest set of Concepts that can categorize all of the documents that match the query, and that represent the content of these documents (i.e. the Concepts and Relations they discuss)" (see Rennison column 26, lines 42-46). In addition, the information space informs "the user about Concepts and Relations between them" (see Rennison column 26, line 62).

Rennison uses "categorization" to automatically categorize documents that match a user's query in its information space. As summarized by Rennison in column 26, lines 63-65, the "problem, therefore, is one of automatic categorization of documents: putting documents in the right categories, and putting subcategories in the

Appl. No. 09/683,235

right categories". Further as Rennison explains in column 25, lines 25-27, unlike "fixed category schemes, the resulting Spaces are dynamically constructed to reflect the Concepts discussed by a specified document set".

In contrast, Appellant categorizes document content to identify a classification label to restrict a query to a category of information in an information retrieval system, where each classification label corresponds to a category of information in the information retrieval system. Appellant's invention as recited in independent claim 1 concerns automatic query formulation, where the formulated query *restricts* a search at an information retrieval system to information concerning a *set of entities* (automatically identified in selected document content) to a *category of information* in the information retrieval system identified by a classification label (assigned by categorizing the selected document content using an organized classification of document content).

B.3 In Summary

Accordingly for the reasons set forth above, Appellant submits that claim 1, reprehensive of group 1, is patentably distinguishable over Rennison. In addition, it should be noted that independent claims 14 and 18 contain the same or very similar limitations to those discussed above with respect to claim 1, and therefore the argument presented above with regard to claim 1 applies equally to independent claims 14 and 18.

Also with regard to dependent claims 3, 5-8, and 10-13 of the first group, these claims depend directly or indirectly from one of independent claims 1 or 14 and thus contain all limitations of the claims from which they depend. Accordingly, the argument presented in this section with regard to independent claims 1 and 14 applies equally to those dependent claims.

C. The Second Group Of Claims (Which Depends From The First Group) Is Patentable Over Rennison

For the purpose of discussion presented herein, claim 2, which depends from claim 1, is discussed as a representative claim of the second group, which includes

Appl. No. 09/683,235

dependent claims 15 and 19, which depend from independent claims 14 and 18, respectively.

Appellant respectfully submits that claim 2 when read as a whole with independent claim 1 is patentably distinguishable over Rennison which provides in addition to the limitations of claim 1 discussed above, the limitation of further limiting the automatically formulated query by adding terms relating to context information surrounding the set of entities in the selected document (i.e., aspect vector 3822 shown in Appellant's Figure 38).

In rejecting claims 2, the Office Action asserts on page 4, first full paragraph that Rennison discloses this aspect of Appellant's claimed limitation in column 21, lines 26-57 and column 26, lines 17-40. Appellant respectfully disagrees.

In column 21, lines 26-57, Rennison discusses algorithms for using information extracted from a document "to generate further Concepts that are good labels for the document" (see Rennison column 21, lines 5-10). These algorithms include algorithms for "query expansion", "co-referencing and weighing", and "deep parsing and summarization" (see Rennison column 21, lines 26-30, lines 31-48, lines 49-57, respectively).

The purpose Rennison identifies additional terms that refer to extracted concepts in a document concerns "Annotation Enhancing" for developing a "series of weighted ConceptIds that are implied topics of the document" (see Rennison column 21, lines 1-4). As set forth in column 19, line 39 to column 20, line 8, Rennison maps concepts extracted from a document to concepts in a knowledge base. The knowledge base serves to constrain the generation of concepts. In effect, the additional terms remove "the dependency upon word choice or morphological inflection of a word referring to a Concept" (see Rennison column 19, lines 58-59).

In contrast, the purpose Appellant identifies additional terms surrounding the set of entities identified in selected document content is to further limit the query automatically formulated which is restricted to a category of information in an information retrieval system identified by an assigned classification label. That is, while Rennison identifies additional terms to improve (i.e., expands the possible) mappings

Appl. No. 09/683,235

between concepts extracted from a document and concepts in a knowledge base, Appellant further constrains the formulation of a query to be applied to a category of information in an information retrieval system.

In column 26, lines 17-40, Rennison discusses two of the three operations in creating an Information space (i.e., graph) of concepts that permits navigation of the set of documents identified with the user query, namely "finding documents that match a user's query" and "organizing the results in a structured space". These aspects of Rennison were discussed above with reference to claim 1. To summarize this discussion, Rennison in "finding documents that match a user's query" performs query *expansion* (see Rennison column 26, lines 25-35), and in "organizing the results in a structured space" categorizes all of the documents that match the user query to build the structured space (see Rennison column 26, lines 42-46).

In contrast, Appellant's claim 2 recites formulating a query by further adding terms defining an assigned classification label. As set forth above, Rennison fails to describe automatically generating a query from selected document content by, in part, (a) *categorizing* the selected document content, and (b) formulating a query to *restrict* a search to a category of information at an information retrieval system, and (c) adding terms to the query made up of an identified set of entities from context information surrounding the set of entities in the selected document content.

Accordingly, for these reasons and for the reasons set forth above regarding independent claim 1, Rennison fails to disclose the limitations set forth in claim 2, which incorporates all limitations of claim 1. In addition, it should be noted that claims 15 and 19 contain the same or very similar limitations to those discussed above with respect to claim 2, and therefore the argument presented above with regard to claim 2 applies equally to claims 15 and 19.

D. The Third Group Of Claims (Which Depends From The Second Group) Is Patentable Over Rennison

For the purpose of discussion presented herein, claim 4, which depends from claim 2, is discussed as a representative claim of the second group, which includes

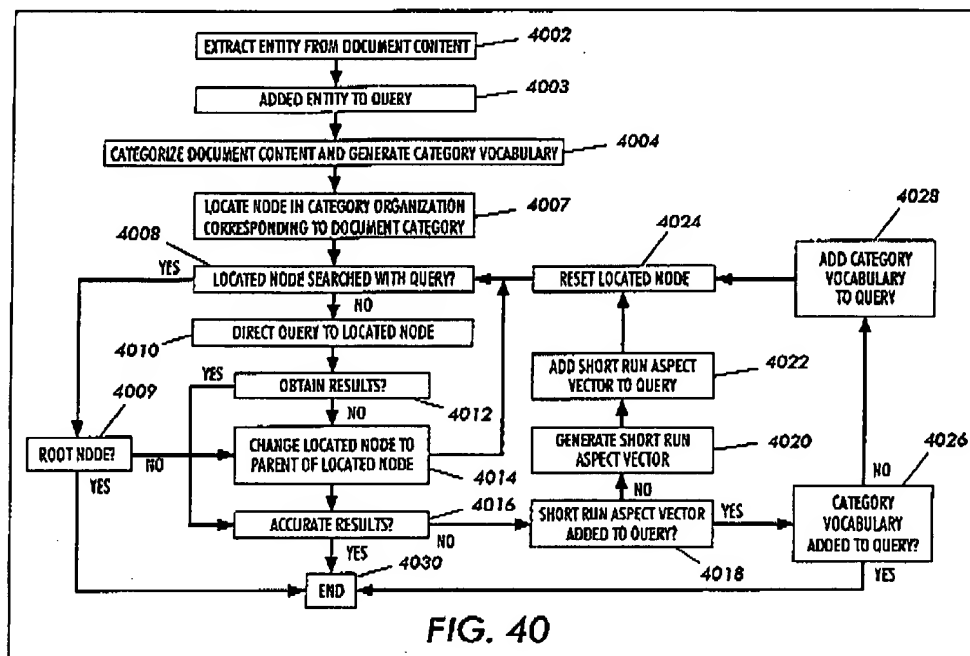
Appl. No. 09/683,235

dependent claims 16 and 20, which depend from depending claims 15 and 19, respectively.

Appellant respectfully submits that claim 4 when read as a whole with independent claim 1 and dependent claim 2 is patentably distinguishable over Rennison which provides in addition to the limitations of claims 1 and 2 discussed above, the limitation of further limiting the automatically formulated query by adding terms defining an assigned classification label (i.e., category vocabulary 3621 shown in Appellant's Figure 38).

In rejecting claim 4 (and 16 and 20), the Office Action asserts on page 4, third full paragraph that Rennison discloses Appellant's claimed limitation in column 21, lines 26-57 and column 26, lines 17-40. Appellant respectfully disagrees.

The cited section of Rennison have been discussed in detail above regarding claim 2, those same arguments are incorporated herein by reference. To illustrate the difference between Appellant's claimed invention recited in claim 4 and the user-navigatable information space described by Rennison, Appellant refers to an example which is described in Appellant's specification in paragraphs 419-425 with reference to Figure 40 reproduced below.



Appl. No. 09/683,235

As illustrated in Appellant's Figure 40, given document content that has been categorized (4004), a node in an ontology is located and searched (4010) with a query (4003) initially defined with an entity extracted from the document content (4002). If accurate results are not identified, first a short run aspect vector is added to the query (4022), and then a category vocabulary is added to the query (4028), before redirecting the search to the located node.

As discussed in detail above with reference to claims 1 and 2, Rennison concerns the creation of a user-navigatable information structure from a large document collection. The user beings by defining a broad query (e.g., "all documents written by Tom Jones from Mar. 1, 1995 to Mar. 1, 1996" – see Rennison column 4, lines 48-50). From the documents collected with the user's query, an information space is created and through which the user may navigate, thereby producing the effect of creating queries and seeing their results (see Rennison column 4, lines 52-61). Rennison in fact differentiates its system by noting that unlike conventional text query systems, information retrieval and query formation are controlled by movement through the information space from one node in the space to another (see Rennison column 12, lines 36-44).

In contrast, Appellant's claim 4 recites formulating a query by further adding terms defining an assigned classification label. As set forth above, Rennison fails to describe either when finding documents that match a user's query or when thereafter organizing the documents in a structure space to automatically generate a query from selected document content by, in part, (a) *categorizing* the selected document content, and (b) formulating a query to *restrict* a search to a category of information at an information retrieval system, (c) adding terms to the query made up of an identified set of entities from context information surrounding the set of entities in the selected document content, and (d) further adding terms to the query defining an assigned classification label identifying the category of information in the information retrieval system.

Accordingly, for these reasons and for the reasons set forth above regarding independent claim 1 and dependent claim 2, Rennison fails to disclose or suggest the

Appl. No. 09/683,235

limitations set forth in claim 4, which incorporates all limitations of claims 1 and 2. In addition, it should be noted that claims 16 and 20 contain the same or very similar limitations to those discussed above with respect to claim 4, and therefore the argument presented above with regard to claim 4 applies equally to claims 16 and 20.

E. Claim 17 (Which Depends From Claim 16 in The Third Group) Is Patentable Over Rennison

Claim 17 which depends from claims 16, 15, and 14 stands on its own for the reasons discussed below. The Office Action alleges on page 6, second paragraph, that claim 17 is obvious in view of Rennison's disclosure set forth in column 4, line 1 to column 5, line 56. Appellant respectfully disagrees.

Appellant's claim 17 recites a content manager for enriching selected document content with results provided from the information retrieval system using the formulated query (see content manager 208 in Appellant's Figure 38). Appellant defines the term "enrich" in paragraph 119 of Appellant's specification to concern the annotation of a document in accordance with a predefined personality.

In column 4, line 1 to column 5, line 56 cited in the Office Action, Rennison describes how a large document collection is segmented for a user into an information space, which provides cues to scale, context, and types of relationships to the user concerning the collection of documents (see Rennison column 3, lines 52-61). Further the cited section describes, as discussed above, how the user may interact with the information space (or information structure) to create queries and see the results of the queries by navigating through the information space (see Rennison column 3, lines 52-61).

However, the sections of Rennison cited in the Office Action fail to disclose or suggest, as recited by Appellant in claim 17, the "enrichment" or "annotation" of document content with search results provided from an information retrieval system using an automatically formulated query. Moreover, as discussed above Rennison further fails to describe or suggest, as recited by Appellant in claim 17 which read together with claims 16, 15, and 14, the automatic formulation of a query that is used to

Appl. No. 09/683,235

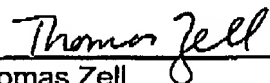
query an information provider for the results that are used to enrich the document content.

Accordingly, Appellant submits that claim 17, which stands on its own, is patentably distinguishable over Rennison for the reasons set forth above and for those reasons set forth above regarding claims 16, 15, and 14.

9. Conclusion

Based on the arguments presented above, claims 1-8 and 10-20 are believed to be in condition for allowance. Appellant therefore respectfully requests that the Board of Patent Appeals and Interferences reconsider this application, reverse in whole the decision of the Examiner, and pass this application for allowance.

Respectfully submitted,



Thomas Zell
Attorney for Appellant
Registration No. 37,481
Telephone: 650-812-4282

Date: April 23, 2004

Appl. No. 09/683,235

APPENDIX A

Claims

1. A method for automatically generating a query from selected document content, comprising:

defining an organized classification of document content with each class in the organized classification of document content having associated therewith a classification label; each classification label corresponding to a category of information in an information retrieval system;

automatically identifying a set of entities in the selected document content for searching additional information related thereto using the information retrieval system;

automatically categorizing the selected document content using the organized classification of document content for assigning the selected document content a classification label from the organized classification of content; and

automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label.

2. The method according to claim 1, further comprising limiting the query by adding terms relating to context information surrounding the set of entities in the selected document content.

3. The method according to claim 2, wherein the number of terms added is limited to a predefined number.

4. The method according to claim 2, further comprising limiting the query by adding terms defining the assigned classification label.

5. The method according to claim 1, wherein the organized classification of document content is defined using a hierarchical organization.

6. The method according to claim 1, further comprising using a text categorizer to assign the classification label assigned from the organized classification of content.

Appl. No. 09/683,235

7. The method according to claim 6, further comprising:

extracting with the text categorizer a set of terms relating to the document content; and

appending to the query ones of the set of terms extracted by the text categorizer to contextualize the query.

8. The method according to claim 7, further comprising abbreviating the set of terms extracted by the text categorizer to a predefined number of terms.

9. The method according to claim 8, wherein said abbreviating comprises:

extracting noun phrases from the selected document content;

ranking the noun phrases by those that occur most frequently in the document content;

defining a subset of noun phrases by identifying those ranked noun phrases that occur more frequently than a first predefined frequency;

ranking those words in the subset of noun phrases by their frequency of occurrence to define an ordered list of words;

defining a subset of the ordered list of words by identifying those ranked words that occur more frequently than a second predefined frequency;

re-ranking the subset of words in inverse frequency to their use in the category of information in the information retrieval system identified by the assigned classification label;

using only those highest ranked words in the re-ranked subset of words to define the set of terms appended to the query.

10. The method according to claim 1, wherein each class in the organized classification of document content has associated therewith a characteristic vocabulary.

11. The method according to claim 10, further comprising ranking results from the query performed at the information retrieval system in accordance with one of the assigned classification label and the characteristic vocabulary.

Appl. No. 09/683,235

12. The method according to claim 11, using the method in a system for enriching selected content of a document with personalities that identify enrichment themes.

13. The method according to claim 1, further comprising automatically identifying the set of entities using a service that recognizes entities of a predefined type.

14. A system for automatically generating a query from selected document content, comprising:

an entity extractor for automatically identifying a set of entities in the selected document content for searching information related thereto using an information retrieval system;

a categorizer for defining an organized classification of document content with each class in the organization of content having associated therewith a classification label; each classification label corresponding to a category of information in the information retrieval system; the categorizer automatically assigning the selected document content a classification label from the organized classification of content; and

a query generator for automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label.

15. The system according to claim 14, further comprising a short length aspect vector generator for generating terms relating to context information surrounding the set of entities in the selected document content; wherein the query generator adds the terms relating to the context information to limit the query.

16. The system according to claim 15, wherein the query generator further limits the query by adding terms defining the selected classification label provided by the categorizer.

17. The system according to claim 16, further comprising a content manager for enriching the selected document content with results provided from the information retrieval system using the query.

Appl. No. 09/683,235

18. An article of manufacture for use in a computer system, comprising:

a memory;

instructions stored in the memory for operating a method for automatically generating a query from selected document content, comprising:

defining an organized classification of document content with each class in the organized classification of document content having associated therewith a classification label; each classification label corresponding to a category of information in an information retrieval system;

automatically identifying a set of entities in the selected document content for searching information related thereto using the information retrieval system;

automatically categorizing the selected document content using the organized classification of document content for assigning the selected document content a classification label from the organized classification of content; and

automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label.

19. The article of manufacture according to claim 18, wherein the instructions stored in the memory further comprise limiting the query by adding terms relating to context information surrounding the set of entities in the selected document content.

20. The article of manufacture according to claim 19, wherein the instructions stored in the memory further comprise further limiting the query by adding terms defining the assigned classification label.